Ecological Indicators 101 (2019) xxx–xxx

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Intensity Analysis and the Figure of Merit's components for assessment of a Cellular Automata – Markov simulation model

Orsolya Gyöngyi Varga[a,*], Robert Gilmore Pontius Jr.[b], Sudhir Kumar Singh[c,1], Szilárd Szabó[a]

[a] Department of Physical Geography and Geoinformatics, Faculty of Science and Technology, University of Debrecen, Debrecen, Hungary
[b] School of Geography, Clark University, Worcester, MA 01610, USA
[c] K. Banerjee Centre of Atmospheric & Ocean Studies, IIDS, Nehru Science Centre, University of Allahabad, Prayagraj 211002, Allahabad, Uttar Pradesh, India

ARTICLE INFO

Keywords:
Cellular automata
CA-Markov
Figure of Merit
Intensity Analysis
Land change
Validation

ABSTRACT

Some popular metrics to evaluate land change simulation models are misleading. Therefore, land change scientists have called for the development of methods to evaluate various aspects of modelling applications. This article answers the call by giving novel methods to compare three types of land change: 1) reference change during the calibration time interval, 2) simulation change during the validation time interval, and 3) reference change during the validation time interval. We compare these changes by using Intensity Analysis' three levels and the Figure of Merit's four components: Misses, Hits, Wrong Hits and False Alarms. We illustrate the concepts by applying a Cellular Automata – Markov land change model to a case study in northeast Hungary. We used reference maps of five land categories to calibrate the model during 2000–2006, then to validate the simulation during 2006–2012. Intensity Analysis' time interval level shows that the simulation change and the reference change decelerated from 2000–2006 to 2006–2012. Intensity Analysis' category level shows that the simulation losses were less than what a pure Markov chain would have dictated. Intensity Analysis' transition level shows that the model's Markov algorithm simulated correctly that the gain of Forest targeted Agriculture and Wetland. The Figure of Merit's components reveals more allocation error than quantity error. Our collection of metrics show that more error derived from the Cellular Automata algorithm than from the Markov algorithm. We recommend that scientists use Intensity Analysis and the Figure of Merit's components to reveal various fundamental aspects of modelling applications.

1. Introduction

Land change models can simulate future changes among land categories (Camacho Olmedo et al., 2018). Use of such models can give insight concerning management options. For example, extrapolation of recent trends can help to anticipate threats to habitats (Bierwagen et al., 2010; Hepinstall et al., 2008; Szabó et al., 2012; Ziółkowska et al., 2014). Proper insight requires that modellers understand how model behavior compares to landscape behavior, which presents several challenges. Therefore, scientists have called for more research into land-change modelling (Paegelow et al. 2013; Pontius et al., 2018; National Research Council 2014). Specifically, Brown et al. (2013) urge that "more needs to be done to develop and disseminate methods for evaluating land-change models". Our article responds to these challenges by presenting methods to compare simulated change to reference change by applying a collection of metrics that give deeper insights than existing popular metrics.

Empirical models typically examine historic patterns of land change during a calibration time interval, and then extrapolate those patterns beyond the calibration time interval. Models simulate temporal change during the extrapolation in terms of two concepts: quantity and allocation. The quantity concerns the size of each transition from one category to another. The allocation concerns the spatial distribution of each transition. Models' algorithms frequently specify the quantity separately from the allocation.

Markov models can describe each transition's quantity. A Markov matrix specifies the proportion of each category that transitions to another category during each time interval. The empirical Markov matrix during the calibration time interval can extrapolate the quantity of each transition beyond the calibration time interval by applying a Markov chain (Baker, 1989). A Markov chain is a popular method of extrapolation in land change models.

* Corresponding author.
E-mail address: varga.orsolya.gyongyi@science.unideb.hu (O.G. Varga).
[1] ORCID: 0000-0001-8465-0649.

https://doi.org/10.1016/j.ecolind.2019.01.057
Received 16 February 2018; Received in revised form 16 November 2018; Accepted 25 January 2019

Cellular Automata (CA) can guide each transition's allocation. CA models consist of a regular grid of cells and rules that dictate how each cell's neighbours influence the future category of each cell (Sipper, 1997). CA models typically simulate transitions in cells that are near the borders between categories. Neumann and Ulam introduced cellular automata in the 1940′s to see whether mathematical formulas and logical rules can describe self-reproduction of biological systems (Benenson and Torrens, 2004).

CA-Markov models combine a Markov algorithm to simulate the quantity of change and a CA algorithm to simulate the allocation of change (Singh et al., 2015). Researchers have applied CA-Markov models to various case studies (Jalerajabi and Ahmadian, 2013; Paegelow et al., 2014; Sang et al., 2011). Some studies compared CA-Markov with other land change models, such as GEOMOD and Idrisi's Land Change Modeller (Camacho Olmedo et al., 2015; Pontius and Malanson, 2005).

CA-Markov is one type of model that simulates transitions among categories, while many others exist. For instance, SLEUTH is a CA model but SLEUTH does not use a Markov matrix to extrapolate the quantity of each transition (Clarke et al., 1997; Silva and Clarke, 2002). SLEUTH has been used for setting up scenarios under various conditions for forecasting urban growth based on historical and contemporary conditions (Herold et al., 2003). Some models are neither CA nor Markov. For example, some models focus on economic factors, where land occupation is based on market conditions, such as in Computable General Equilibrium and Partial Equilibrium models (De Rosa et al., 2016). The structure of land change models vary based on their purposes. Some researchers aim to analyse hotspots of land change at a national level (Verburg et al., 2002) or at spatial resolutions as detailed as the household level (Evans and Kelley, 2004). Some models project changes by analyzing socio-economic and environmental drivers together (Veldkamp and Verburg, 2004). There is a need to integrate model results into landscape planning because environmental management is a typical purpose (Lippe et al., 2017; Convertino and Valverde, 2013). It is useful to know the implications of an extrapolation of recent trends so that decision-makers can understand the trajectory of the system. Regardless of model selection or purpose, modellers should know three aspects of any application: 1 how the model characterizes change during the calibration interval, 2 how the model extrapolates the change during a validation interval, and 3 how the extrapolated change compares to the reference change during the validation interval.

Some scientists compared the model's output map at the final time point of the validation interval to the reference map at the same time point to measure the accuracy of the simulation (Yang et al., 2014; Halmy et al., 2015; Singh et al., 2015, Mishra and Rai, 2016; Keshtkar and Voigt, 2016; Chakraborti et al., 2018). That comparison cannot give insight to temporal change, because both maps show the same time point. Therefore, that two-map comparison cannot distinguish between correctly simulated change and correctly simulated persistence during the validation time interval. If persistence dominates a landscape, then the two-map comparison typically gives large values for percent correct and kappa, regardless of whether the model simulates change correctly. In order to avoid this conceptual blunder, Pontius et al. (2008, 2011, 2018) recommend the use of three maps to compare simulation change versus reference change during the validation time interval. The three maps are: reference at the start of validation interval, simulation at the end of validation interval, and reference at the end of validation interval. The Figure of Merit (FOM) is a popular metric for model validation using this three-map comparison (Klug et al., 1992; Perica and Foufoula-Georgiou, 1996). The FOM ranges from zero to one, where zero means no intersection between simulation and reference change while one means perfect intersection between simulation and reference change. The FOM has limited ability to offer insight because the FOM is a single metric that combines information concerning quantity and allocation. For example, the FOM fails to reveal how the quantity of simulated change compares to the quantity of reference change. Furthermore, FOM fails to show how quantity disagreement compares to allocation disagreement. Our article shows how to compute and interpret FOM's components in a manner that distinguishes between quantity and allocation.

Intensity Analysis can offer insights to modelling applications because Intensity Analysis is a framework to reveal various patterns of change among categories across time intervals (Aldwaik and Pontius, 2012; Aldwaik and Pontius, 2013). Intensity Analysis has three levels, where each increasing level examines increasingly detailed information given the previous level. Intensity Analysis has become popular to analyse temporal changes among categories (Castro and Rocha, 2015; Raphael John et al., 2014; Yang et al., 2017; Quan et al., 2019; Rocha et al., 2017; Aabeyir et al., 2017; Zhou et al., 2014; Huang et al., 2018; Huang et al., 2012). To the best of our knowledge, our article is the first to use Intensity Analysis to evaluate the application of a simulation model.

There are various reasons why the simulation change might not match the reference change during the validation interval. First, the reference change might not be stationary from the calibration interval to the validation interval, in which case empirical calibration would likely produce an extrapolation that lacks predictive power. Second, the model might simulate processes that do not exist in the landscape, such as Markov processes that dictate the quantity of change or neighbourhood processes that dictate the allocation of change. Therefore, proper interpretation requires clear methods to compare three time intervals: 1) reference change during the calibration interval, 2) simulation change during the validation interval, and 3) reference change during the validation interval. Previous methods focused exclusively on the validation interval, which offers helpful but limited insight because such methods fail to consider differences between the calibration and validation intervals. One of the innovations of our article is that we compare the calibration interval to the validation interval.

We illustrate the concepts using a case study in Northeast Hungary. We applied Idrisi Selva's CA-Markov model, and then evaluated the application by using Intensity Analysis and the FOM's components. We compare three time intervals: reference 2000–2006, simulation 2006–2012, and reference 2006–2012. Our objective is to show how Intensity Analysis and FOM's components offer valuable insights concerning how model behavior relates to landscape behavior. The combined use of these measurements and the comparison of the calibration interval to the validation interval are the two main innovations of our paper.

2. Methods

2.1. Study site

The study site is a 25 × 25 km region located around Tokaj city and the tributary of Tisza and Bodrog rivers in Hungary. The site is a diverse landscape of five topographically different microregions (Dövényi, 2010). A large part of the region has been a nature reserve since 1986, and since 2002 has belonged to the Tokaj Wine Region Historic Cultural Landscape, which is a UNESCO World Heritage site (Kerényi, 2015). The site's protected status restricts large land changes.

2.2. Data and simulation

We used maps of the Corine Land Cover (CLC), produced by the European Environment Agency and managed by the Copernicus Land Monitoring Service (https://land.copernicus.eu/pan-european/corine-land-cover). CLC data are popular for landscape monitoring and analysis, including in Hungarian study areas (Csorba and Szabó, 2009; Túri, 2010). Büttner et al. (2004) report the data have a thematic accuracy of at least 85%. The CLC programme established land cover layers via visual interpretation of satellite images at a 1:100,000 scale,

Table 1
Categories in our land change model along with the equivalent CLC category (standard level I) and the content of each category.

Category in model	Category in CLC	Description of category in our study area
Artificial	Artificial surfaces	All urban facilities (including industrial areas) and mining sites
Agriculture	Agricultural	Mainly agricultural areas with various cultures (arable land, vineyards, fruit plantations, pastures, etc.)
Forest	Forests and semi-natural	Mainly broad-leaved and mixed forests with transitional areas into scrub
Wetland	Wetlands	Inland wetlands
Water	Water bodies	All forms of water bodies, including natural and man-made water bodies or rivers.

minimum mapping unit of 25 ha and width of linear objects of 100 m.

CLC categories have three hierarchical levels (Feranec et al., 2016). The most detailed third level has 44 categories, of which 18 appear in our study region. We used the first level, which has five aggregated classes, which we name Artificial, Agriculture, Forest, Wetland, and Water. Table 1 describes our five land cover classes and their equivalent class in CLC nomenclature. CLC data has been distributed in the standard European Coordinate Reference System defined by the European Terrestrial Reference System 1989 (ETRS89) datum and Lambert Azimuthal Equal Area projection (EPSG: 3035). We obtained vector maps at 2000, 2006, and 2012, and then converted them into 25 m spatial resolution raster layers in the software Idrisi Selva.

We used the change during 2000–2006 to calibrate the CA-Markov model. The model then simulated changes during 2006–2012, which is the validation time interval. The CA-Markov model has distinct algorithms to simulate the quantity versus the allocation of each transition.

The model's Markov algorithm guides the simulation's quantity. The algorithm computes a Markov matrix based on the changes during the calibration time interval, and then uses a Markov chain to extrapolate the size of each transition during subsequent time intervals. The Markov chain assumes a constant proportion of each initial category transitions to every other category during each time interval.

The model's CA algorithm guides the simulation's allocation. The algorithm allows the simulation of a spatial process whereby each category gains near the edges of the category's initial patches (Eastman, 2012; Mas et al., 2014). A spatial filter and an iteration number influence how near to the edges the changes occur. We used a 5-by-5 spatial filter and an iteration number of six, which are the model's default parameters. Idrisi Selva's CA-Markov model does not have automated calibration for these two parameters.

Fig. 1 shows the maps that serve as the basis of our analysis. At all three time points, Artificial accounts 5–6% of the spatial extent, Agriculture for 72–74%, Forest for 14–16%, Wetland for 4% and Water for 3%.

2.3. Intensity Analysis

Intensity Analysis is a framework to understand the sizes and intensities of temporal changes among categories (Aldwaik and Pontius, 2012, 2013; Pontius et al., 2013). Intensity Analysis has three levels: Interval, Category, and Transition. The Interval level examines the overall change during each time interval. The Category level examines the loss and gain of each category during each time interval. The Transition level examines how the gain of a category transitions from other categories during each time interval. We applied Intensity Analysis using free software from http://www.clarku.edu/~rpontius/. The inputs were a crosstabulation matrix for each of three time intervals: 2000–2006 reference, 2006–2012 simulation, and 2006–2012 reference.

Table 2 gives the mathematical notation for the equations of Intensity Analysis based on Pontius et al. (2013). All time intervals have the same duration of six years; therefore, we did not use the equations of Aldwaik and Pontius (2012), which compute annual change during time intervals that have various durations.

For the interval level, Eq. (1) defines S_t, which is the change percentage during each interval t. The change percentage S_t is the uniform intensity during interval t for the category level. Eqs. (2) and (3) give the category level intensities of loss L_{ti} and gain G_{tj} during interval t. If change during interval t were distributed uniformly across the spatial extent, then $S_t = L_{ti} = G_{tj}$ for all categories i and j. If $L_{ti} < S_t$, then the loss of category i is dormant during interval t. If $L_{ti} > S_t$, then the loss of category i is active during interval t. Similarly, if $G_{tj} < S_t$, then G_{tj} is dormant; and if $G_{tj} > S_t$, then G_{tj} is active. If the status as dormant or active is the same during sequential time intervals, then we say the category's loss or gain is stationary. The loss intensity L_{ti} is identical to the diagonal entry in a Markov matrix for interval t concerning category i.

$$S_t = \frac{(\text{size of change during interval } t)100\%}{\text{size of spatial extent}} = \frac{\left\{\sum_{i=1}^{J}\left[\left(\sum_{j=1}^{J} C_{tij}\right) - C_{tii}\right]\right\}100\%}{\sum_{i=1}^{J}\left(\sum_{j=1}^{J} C_{tij}\right)} \tag{1}$$

$$L_{ti} = \frac{(\text{size of loss of } i \text{ during interval } t)100\%}{\text{size of } i \text{ at start of interval } t} = \frac{\left[\left(\sum_{j=1}^{J} C_{tij}\right) - C_{tii}\right]100\%}{\sum_{j=1}^{J} C_{tij}} \tag{2}$$

$$G_{tj} = \frac{(\text{size of gain of } j \text{ during interval } t)100\%}{\text{size of } j \text{ at end of interval } t} = \frac{[(\sum_{i=1}^{J} C_{tij}) - C_{tjj}]100\%}{\sum_{i=1}^{J} C_{tij}} \tag{3}$$

For the transition level, Eq. (4) gives the transition intensity R_{tij} from category i to category j during time interval t. Eq. (5) gives the uniform intensity W_{tj} for the gain of category j from categories that are not j at the interval's start time. The order of subscripts j and i in C_{tji} in the denominator of Eq. (5) is intentional, so that the summation over i subtracts category j at the interval's start time. If category j were to gain uniformly from all other categories, then $W_{tj} = R_{tij}$ for all i. If $R_{tij} < W_{tj}$, then the gain of j avoids i. If $R_{tij} > W_{tj}$, then the gain of j targets i. If the status as avoiding or targeting is the same during sequential time intervals, then we say the transition is stationary. The transition intensity R_{tij} is identical to the off-diagonal entry in a Markov matrix for interval t concerning the transition from category i to category j.

$$R_{tij} = \frac{(\text{size of transtion from } i \text{ to } j \text{ during interval } t)100\%}{\text{size of } i \text{ at start of interval } t} = \frac{(C_{tij})100\%}{\sum_{j=1}^{J} C_{tij}} \tag{4}$$

$$W_{tj} = \frac{(\text{size of gain of } j \text{ during interval } t)100\%}{\text{size of not } j \text{ at start of interval } t} = \frac{[(\sum_{i=1}^{J} C_{tij}) - C_{tjj}]100\%}{\sum_{i=1}^{J}\left[\left(\sum_{j=1}^{J} C_{tij}\right) - C_{tji}\right]} \tag{5}$$

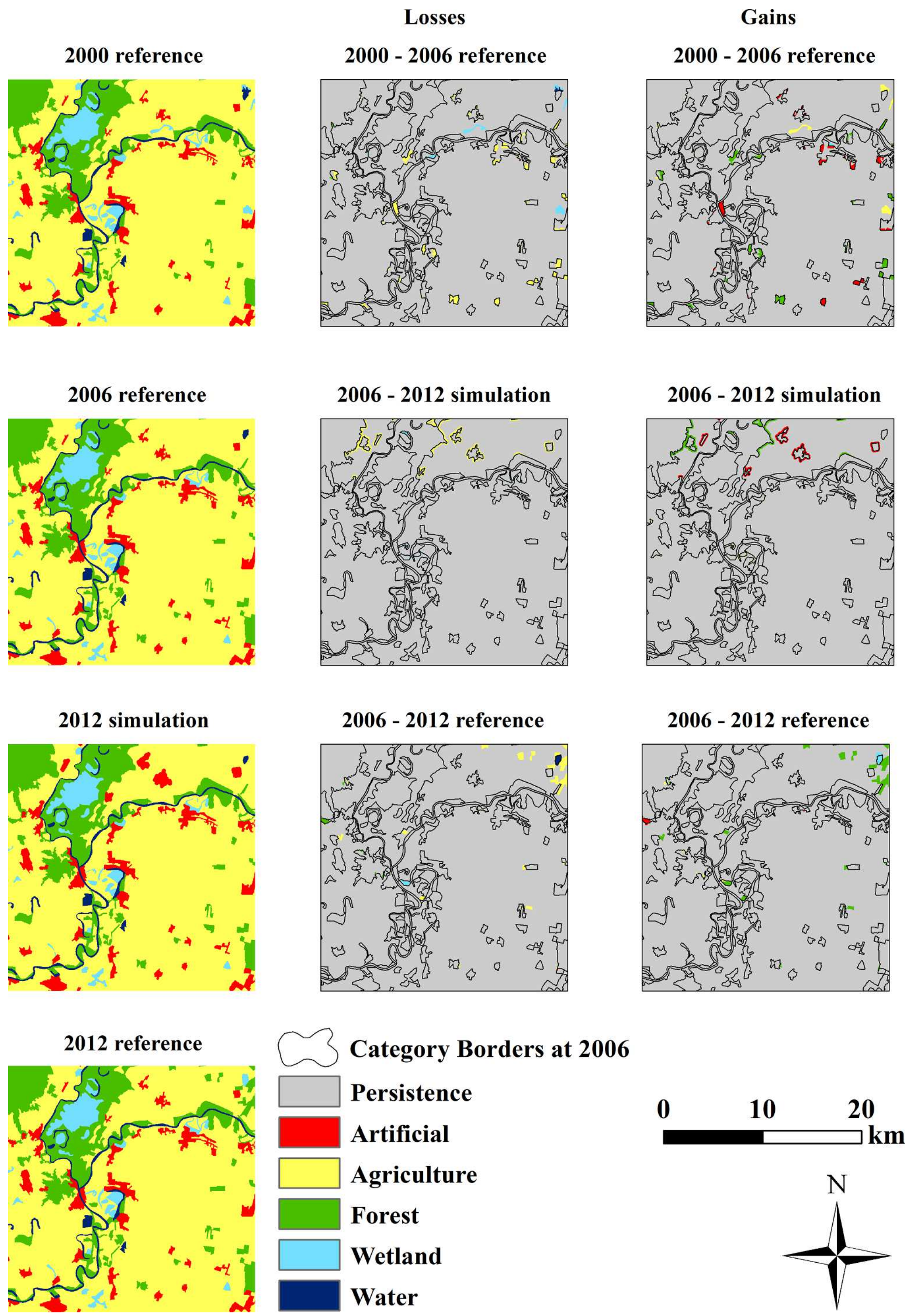

Fig. 1. Reference maps of categories at three time points and of change during two time intervals in northeast Hungary. Persistence means a category remains the same during a time interval.

2.4. Figure of Merit's components

We compare the simulation change to the reference change during 2006–2012 to gain insight concerning model validation. We perform the comparison visually by overlaying three maps: reference 2006, simulation 2012, and reference 2012. We also perform the comparison quantitatively by computing the components of the FOM. The FOM is a ratio, where the numerator is the intersection of simulated and reference change, while the denominator is the union of simulated and reference change. We used the 'lulcc package' of the R 3.3.3 software to compute the FOM's components (Moulds and Buytaert, 2015; R Core Team 2017). Eq. (6) shows how the FOM derives from its four components: Misses, Hits, Wrong Hits and False Alarms (Pontius et al., 2011).

$$Figure\ \ of\ \ Merit = \frac{(Hits)100\%}{Misses + Hits + Wrong\ \ Hits + False\ \ Alarms} \tag{6}$$

where Misses = area of error due to reference change simulated as persistence; Hits = area of correct due to reference change simulated as change; Wrong Hits = area of error due to reference change simulated as change to the wrong category; False Alarms = area of error due to reference persistence simulated as change.

Table 2
Mathematical Notation for Intensity Analysis.

Symbol	Meaning
C_{tij}	number of cells that are category *i* at start and category *j* at end of time interval *t*
C_{tji}	number of cells that are category *j* at start and category *i* at end of time interval *t*
G_{tj}	intensity of gain of category *j* during time interval *t* relative to size of category *j* at end of time interval *t*
i	index for a category
j	index for a category
J	number of categories
L_{ti}	intensity of loss of category *i* during time interval *t* relative to size of category *i* at start of time interval *t*
R_{tij}	intensity of transition from category *i* to category *j* during time interval *t* relative to size of category *i* at start of time interval *t*
S_t	change percentage during time interval *t*
t	index for a time interval
W_{tj}	uniform intensity of transition from all non-*j* categories to category *j* during time interval *t* relative to size of all non-*j* categories at start of time interval *t*

FOM's components allow computation of quantity disagreement, allocation disagreement and total disagreement (Chen and Pontius, 2010; Liu et al. 2014). Eq. (7) gives quantity disagreement, while Eq. (8) gives allocation disagreement. Eq. (9) shows that the total disagreement is the sum of the quantity disagreement, allocation disagreement and Wrong Hits. Wrong Hits are disagreement in the detailed respect that Wrong Hits are places where the simulation map does not match the reference map at the end time of the validation interval. Wrong Hits are agreement in the broad respect that Wrong Hits are places where change occurs according to both the simulation and the reference maps during the validation interval.

$$\textit{Quantity disagreement} = |\textit{False Alarms} - \textit{Misses}| \tag{7}$$

$$\textit{Allocation disagreement} = 2\ \mathrm{MINIMUM}(\textit{False Alarms}, \textit{Misses}) \tag{8}$$

$$\textit{Total disagreement} = \textit{Quantity disagreement} + \textit{Allocation disagreement} + \textit{Wrong Hits} \tag{9}$$

3. Results

3.1. Intensity Analysis

Table 3 shows the number of cells of each transition during three

Table 3
Number of cells that transition from each row's start time category to each column's end time category. For each transition, the top number gives 2000–2006 reference; the middle number gives 2006–2012 simulation; and the bottom number gives 2006–2012 reference. Persistence is a transition from a category to itself. Loss is the row's sum minus persistence. Gain is the column's sum minus persistence. Overall change is in the lower right.

	End Time					
Start Time	Artificial	Agriculture	Forest	Wetland	Water	Loss
Artificial	50,107	44	0	1	0	45
	55,837	0	0	0	0	0
	55,762	9	66	0	0	75
Agriculture	5,730	728,057	7,092	270	5	13,097
	5,618	719,515	6,564	0	0	12,182
	141	722,944	8,256	355	1	8,753
Forest	0	415	140,508	3	5	423
	0	0	147,959	0	2	2
	550	436	146,969	3	3	992
Wetland	0	2,812	359	36,770	0	3,171
	0	2,390	325	34,329	0	2,715
	0	145	528	36,371	0	673
Water	0	369	2	0	27,451	371
	0	361	0	0	27,100	361
	1	2	8	711	26,739	722
Gain	5,730	3,640	7,453	274	10	17,107
	5,618	2,751	6,889	0	2	15,260
	692	592	8,858	1,069	4	11,215

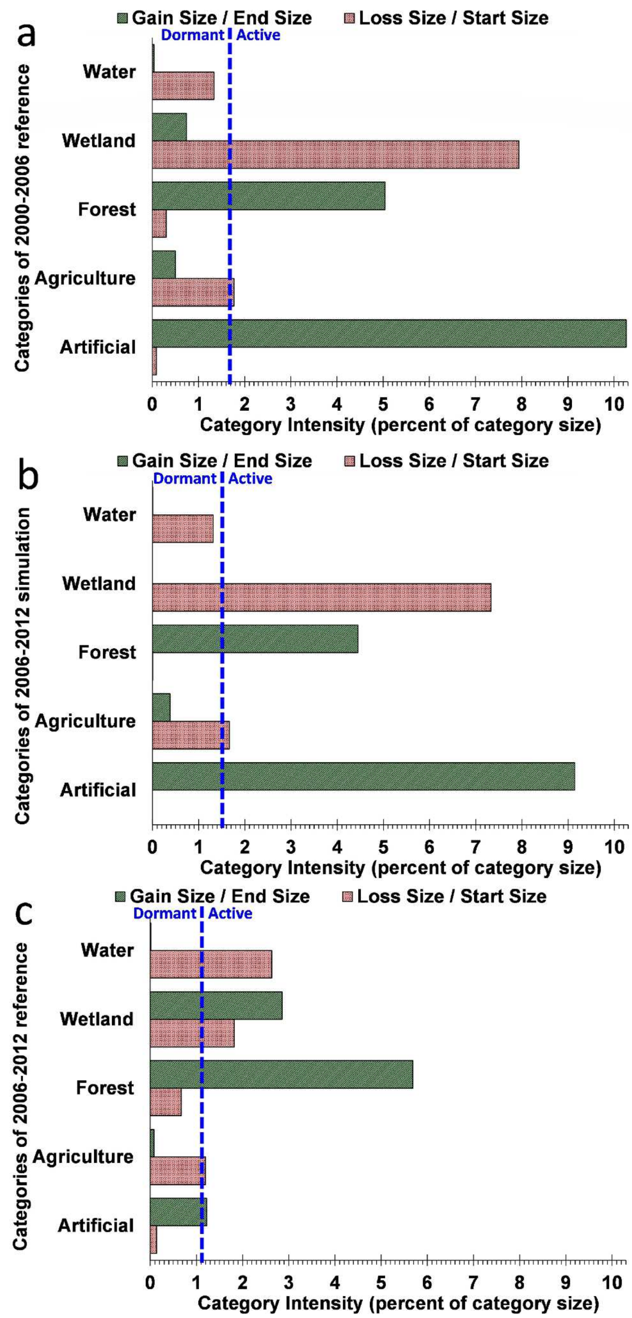

Fig. 2. Category intensities during three time intervals: (a) 2000–2006 reference, (b) 2006–2012 simulation, and (c) 2006–2012 reference.

time intervals. One million cells exist the spatial extent; therefore, each entry in Table 3 divided by ten thousand gives the percentage of the spatial extent. The lower right entry shows that overall reference change during 2000–2006 is 17,107 cells, implying 1.7% of the spatial extent. Overall simulation change during 2006–2012 is 1.5%, while overall reference change during 2006–2012 is 1.1%.

Fig. 2 shows results from Intensity Analysis' interval and category levels. The uniform lines in each graph indicate the interval level in terms of overall change as a percentage of the spatial extent. The model simulated deceleration of overall change from 2000–2006 to 2006–2012, while the simulation deceleration was not as severe as the reference deceleration. Fig. 2a and b show that the dormant or active status of each loss and gain during 2000–2006 was the same as during the simulation. If the software were to have simulated the sizes of the transitions by using a Markov matrix exclusively, then the 2006–2012 simulation loss intensities would be equal to the 2000–2006 reference loss intensities. However, Fig. 2a and b show that the 2006–2012 simulation loss intensities are less than the 2000–2006 reference loss intensities. Fig. 2a and c show that the reference patterns are not stationary from the calibration interval to the validation interval. Most noteworthy, Wetland lost and Artificial gained substantially during the calibration interval but not during the validation interval. Therefore, the categorical intensities during the simulation do not match the reference during 2006–2012. Table 3 shows that Agriculture had the largest size of loss, but Fig. 2 shows that Agriculture did not have the largest intensity of loss during all time intervals. The loss intensity for Agriculture was less than for Wetland because of Agriculture's large size, which is in the denominator of the intensity. Wetland had the greatest loss intensity during the calibration interval and the simulation, which was due in part to Wetland's small size in the denominator of the intensity.

Fig. 3 shows results of the transition level Intensity Analysis for the two largest gains: Artificial and Forest. Comparison of the 2000–2006 reference and the simulation show how the CA-Markov model extrapolated the intensity of changes from the calibration interval to the validation interval. If the software were to have simulated the sizes of the transitions by using a Markov matrix exclusively, then the 2006–2012 simulation transition intensities would be identical to the 2000–2006 reference transition intensities. The gain of Artificial is not stationary through time. The gain of Artificial targeted only Agriculture during the calibration interval and the simulation. However, the reference gain of Artificial targeted only Forest during the validation interval. In contrast, the gain of Forest is stationary across the three intervals with respect to how the gain of Forest avoided or targeted the non-Forest categories.

3.2. Figure of Merit's components

Fig. 4 shows the Figure of Merit's components. The 2006–2012 reference change is the union of Misses, Hits, and Wrong Hits. The 2006–2012 simulation change is the union of Hits, Wrong Hits and False Alarms. The CA-Markov model allocated the gain of each category around patches of the category at 2006, which caused long winding patches of simulation change. The shapes of the patches of simulation change do not match the compact and isolated patches of reference change. Correctly simulated persistence accounts for 97% of the spatial extent, which is why overall percent correct and kappa at the end time point are misleading measurements of a model's ability to simulate change.

The Figure of Merit is 0.07%, which is the size of Hits as a percentage of the sum of sizes of the four components. Fig. 5 shows that Hits accounted for 0.02% of the spatial extent. Reference change during 2006–2012 accounted for 1.12% of the spatial extent, which is the sum of Misses, Hits, and Wrong Hits. Simulation change accounted for 1.53% of the spatial extent, which is the sum of Hits, Wrong Hits and False Alarms. Quantity disagreement is 0.41% while allocation disagreement is 2.12% of the spatial extent. Total disagreement is 2.57%, which is the sum of Misses, Wrong Hits and False Alarms.

4. Discussion

4.1. Quantity disagreement and Intensity Analysis

The CA-Markov model uses a Markov chain to guide the simulation's quantity of each transition. Intensity Analysis shows how the simulation produced small differences with respect to 2006–2012 reference concerning quantity.

Intensity Analysis' interval level showed that the model simulated correctly the deceleration of overall change from the calibration interval to the validation interval. Many Markov chains lead to a steady state concerning the size of each category, in which case the Markov chain extrapolates a deceleration of change. Intensity Analysis' category level showed that the Markov algorithm simulated the dormant or active status of each category's loss and gain as the category's same status during the calibration interval. Intensity Analysis' transition level showed that the simulated gain of Artificial targeted Agriculture, as was the case during the calibration interval; however, the reference gain of Artificial targeted Forest during the validation interval. The simulation did not match the reference pattern during the validation interval because the reference pattern was not stationary concerning transitions to Artificial. Intensity Analysis' transition level showed that the simulated gain of Forest targeted Agriculture and Wetland, which is a pattern that was stationary through time according to the reference data.

Additional analysis showed that Idrisi Selva's CA-Markov model simulated fewer and smaller transitions than an extrapolation of a Markov chain would dictate. Table 3 shows that Artificial experienced loss and Wetland experienced gain during the calibration interval, but the CA-Markov model simulated zero loss of Artificial and zero gain of Wetland. This illustrates how the CA-Markov did not follow the quantities that a pure Markov extrapolation would have dictated.

4.2. Allocation disagreement and Figure of Merit's components

The CA-Markov model uses the Cellular Automata algorithm primarily to guide the change's allocation. FOM's components showed how the simulation had substantial differences related to the 2006–2012 reference concerning allocation.

Hits and Wrong Hits were near zero, which indicates that the simulation change did not correspond to the reference change. If Misses or False Alarms equal zero, then allocation difference is zero. If Misses equal False Alarms, then quantity difference is zero. If Misses are greater than False Alarms, then reference change is greater than simulated change. If Misses are less than False Alarms, then reference change is less than simulated change, which our case study illustrates. In our application, the sizes of Misses and False Alarms imply that allocation difference was greater than quantity difference. This implies that disagreement during the validation interval derived more from the model's Cellular Automata algorithm than from its Markov algorithm.

If we had examined only the single FOM metric, then we would not be able to have the insights that we had from interpretation of Misses, Hits, Wrong Hits and False Alarms. FOM's single number combines quantity disagreement and allocation disagreement into one measurement, which fails to reveal whether disagreement derives from quantity or allocation. If the quantity of simulation change differs substantially from the quantity of reference change during the validation interval, then it is possible for FOM to be small, even when the simulation allocates change as accurately as possible. For example, if the simulation change is a small subset of the reference change, then then FOM will be small, even when False Alarms are zero. If the reference change is a small subset of the simulation change, then the FOM will be small, even when Misses are zero. FOM's components reveal the reasons for the size of the FOM.

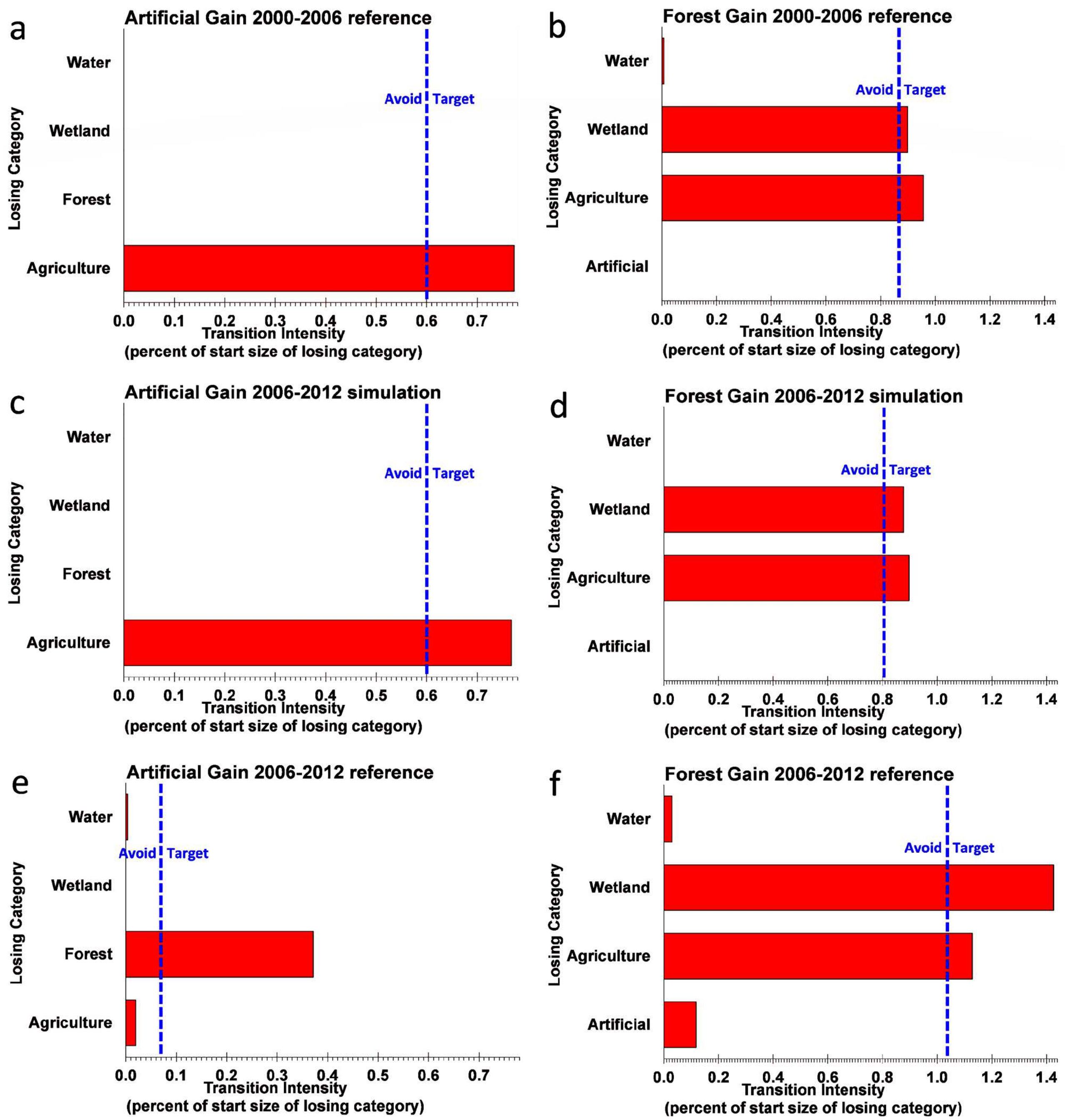

Fig. 3. Transition intensities for the gains of Artificial and Forest during three time intervals: (a-b) 2000–2006 reference, (c-d) 2006–2012 simulation, and (e-f) 2006–2012 reference.

In our application, the Cellular Automata algorithm did not use the 2000–2006 reference change to calibrate the allocation of simulated change. The algorithm's spatial filter causes a category to gain around the edges of the category's patches. However, Fig. 1 shows that reference change in our study area is rarely allocated around the edges of patches. Our use of the spatial filter may be one reason for the large allocation disagreement.

We performed sensitivity analysis to see how the size of the spatial filter influences the results. We ran the model with spatial filters of 3-by-3, 5-by-5 and 7-by-7. Output showed trivial variation in the simulation. The variation of the spatial filter caused variation in the quantity of change for at most 36 cells of simulation loss of Wetland. Figure of Merit was 1.6% for 3-by-3, 0.7% for 5-by-5 and 0.6% for 7-by-7. All three sizes of the spatial filter cause simulated change to occur near patch edges, whereas a larger spatial filter allows change to occur slightly farther from patch edges. The sensitivity results suggest that reference change is slightly more concentrated directly adjacent to patch edges in the rare cases where reference change exists near patch edges. Increase of the iteration parameter from six to twelve caused FOM values to shrink. More sophisticated sensitivity analysis for model parameters is a topic for future research (Saltelli et al., 2008).

Some investigators have compared CA-Markov model runs that used a spatial filter to runs that did not use a spatial filter (Camacho Olmedo et al., 2015; Pontius and Malanson, 2005). They found that the spatial filter influenced the simulation's allocation, but did not influence Hits substantially.

Another possible reason for the large allocation disagreement might be that we did not use suitability maps to guide the spatial allocation.

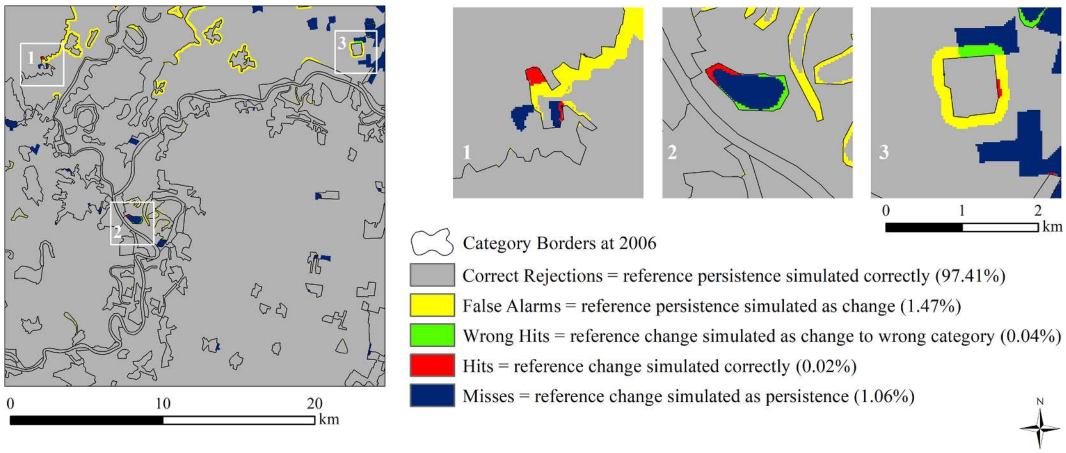

Fig. 4. Three-map comparison to examine simulation versus reference change during 2006–2012. The numbered boxes show three regions that contain Hits.

Idrisi's CA-Markov allows inclusion of suitability maps that use independent variables for calibration. We did not use such suitability maps because our purpose was to show methods for model assessment. Other authors used suitability maps in their applications, while they saw results similar to ours concerning allocation disagreement (Memarian et al., 2012; Pontius et al., 2008; Pontius et al., 2011). Even if we were to have used suitability maps, the CA's spatial filter would have still caused a category to gain around the category's existing patches. If the reference maps do not show spatial dependency and the goal is predictive power, then the modeler should not use a spatial filter.

Some modellers are tempted to modify the simulation model in an effort to increase accuracy. A modeller should first have a specific goal for a particular validation metric before modifying the model. The goal will help the modeler to decide where to focus attention. Deep thought is necessary to select a relevant validation metric and a goal for the metric. The modeller must consider the particular applied research question to select the metric and its goal. In our application to Hungary, the size of simulation change was 1.53% of the spatial extent and the reference change was 1.12%. If the main goal is to simulate the quantity of change, then perhaps the simulation of somewhat more than the reference change is tolerable, while allocation difference is less important. For example, if the goal is to simulate disturbance of carbon in a region where carbon density is spatially uniform, then quantity difference determines error of carbon disturbance, while allocation difference is irrelevant (Pontius, 2018). However, if carbon density is not spatially uniform, then allocation difference can be important for simulation of carbon disturbance. Modellers must consider the goal of the simulation before jumping to an endless chase to increase accuracy. This article gives metrics to help modelers align the goal of the simulation with various aspects of the model. For this article's Hungarian example, validation results showed allocation disagreement is much larger than quantity disagreement. So if the goal is to decrease total disagreement, then the modeler should focus on the allocation of change. The first step would be to simplify the CA-Markov model by eliminating the spatial filter, because the reference change is not concentrated near patch edges. The second step would be to use suitability maps to guide the allocation of simulated change. Idrisi's CA-Markov model has the ability to include such suitability maps.

4.3. Limitations and pitfalls of popular metrics

Some scientists aim to use a single metric to evaluate modelling applications. However, any single metric cannot offer insights concerning various aspects of modelling applications. For example, a popular and misleading metric is the percent correct between the simulation map and the reference map at the end time point of the validation interval (Kityuttachai et al., 2013). Our application was 97% correct according to a two-map comparison between the simulation and the reference maps at 2012. Persistence simulated correctly is the reason for the large percent correct. Percent correct at the validation interval's end time point fails to distinguish between correctly simulated persistence versus correctly simulated change. Clear interpretation is limited or impossible for some other popular metrics, such as FOM and the kappa index of agreement (Yang et al., 2014; Subedi et al., 2013; Parsa et al., 2016; Chakraborti et al., 2018). Scientists claim

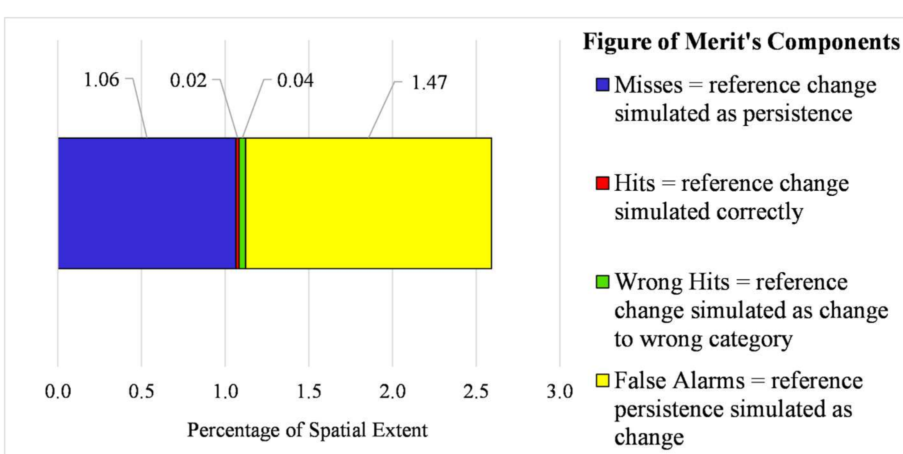

Fig. 5. Figure of Merit's components as percentages of the spatial extent.

kappa is an index that accounts for random agreement. But kappa accounts for randomness in a confusing, misleading and irrelevant manner (Pontius and Millones, 2011). Furthermore, kappa is not appropriate for validation of temporal change because kappa compares two maps at a single time point, thus cannot give insight concerning temporal change. The FOM examines temporal change during the validation interval, but the FOM offers limited interpretation because the FOM combines quantity disagreement and allocation disagreement into a single metric. A metric that combines various concepts can be difficult to interpret (Bradley et al., 2016). It is more helpful to use a collection of metrics, where each metric reveals a distinct and clear aspect of the modelling application. Furthermore, we recommend authors show maps that reveal reference change during the calibration interval, simulation change during the validation interval, and reference change during the validation interval. An overlay of the latter two maps show Misses, Hits, Wrong Hits, False Alarms and Correct Rejections, which communicates clearly the quantity and allocation of changes during the validation interval (Shafizadeh-Moghadam et al., 2017).

5. Conclusions

We have presented novel methods to interpret applications of land change models. Our collection of metrics reveals various aspects that are helpful to understand simulations of temporal change. For our CA-Markov modelling application for a Hungarian case study, Intensity Analysis' interval level shows the model simulated more change than the reference change during the validation interval, because the reference change decelerated from the calibration interval to the validation interval. Intensity Analysis' category level shows the CA-Markov model did not follow exactly the loss intensities that a pure Markov chain would imply. Intensity Analysis' transition level shows the model simulated correctly that the gain of Forest targeted Agriculture and Wetland. Hits were almost zero, which indicates almost no intersection between simulated and reference change during the validation interval. Misses and False Alarms showed that allocation difference was larger than quantity difference, which reflects how the Cellular Automata algorithm caused more error than the Markov algorithm.

We conclude with recommendations that apply generally. Scientists must compare visually and quantitatively the changes during three intervals: (i) reference change during the calibration interval, (ii) simulation change during the validation interval, and (iii) reference change during the validation interval. Comparison between (i) and (ii) relates the calibration patterns to the subsequent simulation. Comparison between (ii) and (iii) distinguishes between simulation and reference changes during the validation interval. Comparison between (i) and (iii) shows the degree to which the reference patterns are stationary through time. For each comparison, Intensity Analysis reveals various levels of information concerning quantity disagreement. The FOM's components distinguish quantity disagreement from allocation disagreement during the validation interval. Our recommended collection of metrics generate insights that are deeper than any single metric can communicate.

Conflicts of interest

The authors have no conflicts of interest.

Acknowledgements

The United States National Science Foundation (NSF) supported this work via the Long Term Ecological Research network via grant OCE-1637630 for Plum Island Ecosystems. On behalf of Dr. Szilárd Szabó, our article was supported by the EFOP-3.6.1-16-2016-00022 and the TNN 123457 projects. The projects were co-financed by the European Union and the European Social Fund. Anonymous reviewers supplied constructive feedback that helped to improve this article.

References

Aabeyir, R., Agyare, W.A., Weir, M.J.C., et al., 2017. Multi-level land cover change analysis in the forest-Savannah transition zone of the Kintampo Municipality, Ghana. J. Nat. Resourc. Develop. 7, 1–11.
Aldwaik, S.Z., Pontius Jr., R.G., 2012. Intensity analysis to unify measurements of size and stationarity of land changes by interval, category, and transition. Landscape Urban Plann. 106, 103–114.
Aldwaik, S.Z., Pontius Jr., R.G., 2013. Map errors that could account for deviations from a uniform intensity of land change. Int. J. Geogr. Inform. Sci. 27, 1717–1739.
Baker, W., 1989. A review of models of landscape change. Landscape Ecol. 2, 111–133.
Benenson, I., Torrens, P.M., 2004. Geosimulation: Automata-Based Modeling of Urban Phenomena. John Wiley Sons Ltd.
Bierwagen, B.G., Theobald, D.M., Pykec, C.R., et al., 2010. National housing and impervious surface scenarios for integrated climate impact assessments. PNAS 107 (49), 20887–20892.
Bradley, A.V., Rosa, I.M.D., Pontius Jr., R.G., et al., 2016. SimiVal, a multi-criteria map comparison tool for land-change model projections. Environ. Modell. Software 82, 229–240.
Brown, D.G., Verburg, P.H., Pontius Jr., R.G., et al., 2013. Opportunities to improve impact, integration, and evaluation of land change models. Curr. Opin. Environ. Sustain. 5, 452–457.
Büttner, G., Feranec, J., Jaffrain, G., et al., 2004. The CORINE land cover 2000 project. EARSeL eProceedings.
Camacho Olmedo, M.T., Paegelow, M., Mas, J., Escobar, F., 2018. Geomatic Approaches for Modeling Land Change Scenarios. In: Cartwright, W., Gartner, G., Meng, L., Peterson, M.P. (Eds.), Lecture Notes in Geoinformation and Cartography. Springer International Publishing AG, Cham, Switzerland.
Camacho Olmedo, M.T., Pontius Jr., R.G., Paegelow, M., et al., 2015. Comparison of simulation models in terms of quantity and allocation of land change. Environ. Modell. Software 69, 214–221.
Castro, G.G.H., Rocha, W.P., 2015. Change analysis of land use and urban growth in the municipalities of Culiacan and Navolato, Sinaloa, Mexico using statistical techniques and GIS. J. Geogr. Inform. Syst. 7 (6), 620–630.
Chakraborti, S., Das, D.N., Mondal, B., et al., 2018. A neural network and landscape metrics to propose a flexible urban growth boundary: a case study. Ecol. Ind. 93, 952–965. https://doi.org/10.1016/j.ecolind.2018.05.036.
Chen, H., Pontius Jr., R.G., 2010. Diagnostic tools to evaluate a spatial land change projection along a gradient of an explanatory variable. Landscape Ecol. 25, 1319–1331.
Clarke, K.C., Hoppen, S., Gaydos, L., 1997. A self-modelling cellular automata model of historical urbanization in the San Francisco bay area. Environ. Plan. 24, 247–261.
Convertino, M., Valverde Jr., J., 2013. Portfolio decision analysis framework for value-focused ecosystem management. PLOS One 9 (3), e92951.
Csorba, P., Szabó, S., 2009. Determination of human transformation of landscapes: a case study from Hungary. Hungarian Geogr. Bull. 58, 91–99.
De Rosa, M., Knudsen, M.T., Hermansen, J.E., 2016. A comparison of Land Use Change models: challenges and future developments. J. Cleaner Prod. 113, 183–193.
Dövényi, Z., 2010. Inventory of Microregions in Hungary. MTA Földrajztudományi Kutatóintézet, Budapest.
Eastman, J.R., 2012. Idrisi Selva Manual. Clark University, Worcester MA.
Evans, T.P., Kelley, H., 2004. Multi-scale analysis of a household level agent-based model of landcover change. J. Environ. Manage. 72, 57–72.
Feranec, J., Soukup, T., Hazeu, G., et al., 2016. European Landscape Dynamics: CORINE Land Cover. Data. CRC Press.
Halmy, M.W.A., Gessler, P.E., Hicke, J.A., et al., 2015. Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. Appl. Geogr. 63, 101–112.
Hepinstall, J.A., Alberti, M., Marzluff, J.M., 2008. Predicting land cover change and avian community responses in rapidly urbanizing environments. Landscape Ecol. 23 (10), 1257–1276.
Herold, M., Goldstein, N.C., Clarke, K.C., 2003. The spatiotemporal form of urban growth: measurement, analysis and modeling. Rem. Sens. Environ. 86, 286–302.
Huang, B., Huang, J., Pontius Jr., R.G., Tu, Z., 2018. Comparison of Intensity Analysis and the land use dynamic degrees to measure land changes outside versus inside the coastal zone of Longhai, China. Ecol. Indic. 89, 336–347.
Huang, J., Pontius Jr., R.G., Li, Q., Zhang, Y., 2012. Use of intensity analysis to link patterns with processes of land change from 1986 to 2007 in a coastal watershed of southeast China. Appl. Geogr. 34, 371–384.
Jalerajabi, P., Ahmadian, R., 2013. Simulation of future land use/cover changes in Zanjan City, Iran based on the CA-Markov model and the assessment of its impact on surroundings rural settlements. Int. J. Comput. Theory Eng. 5, 726–730.
Kerényi, A., 2015. Loess Features on Tokaj Hill. In: Lóczy, D. (Ed.), Landscapes and Landforms of Hungary, 1st ed. Springer International Publishing, pp. 219–225.
Keshtkar, H., Voigt, W., 2016. A spatiotemporal analysis of landscape change using an integrated Markov chain and cellular automata models. Model. Earth Syst. Environ. 2, 10.
Kityuttachai, K., Tripathi, N.K., Tipdecho, T., et al., 2013. CA-Markov analysis of constrained coastal urban growth modeling: Hua Hin Seaside City, Thailand. Sustainability 5, 1480–1500.
Klug, W., Graziani, G., Grippa, G., et al., 1992. Evaluation of Long Range Atmospheric Transport Models using Environmental Radioactivity Data from the Chernobyl Accident: The ATMES Report.
Lippe, M., Hilger, T., Sudchalee, S., Wechpibal, N., Jintrawet, A., Cadisch, G., 2017. Simulating stakeholder-based land-use change scenarios and their implication on

above-ground carbon and environmental management in Northern Thailand. Land 6 (4), 85.

Liu, Y., Feng, Y., Pontius Jr., R.G., 2014. Spatially-explicit simulation of urban growth through self-adaptive genetic algorithm and cellular automata modelling. Land 3 (3), 719–738.

Mas, J., Kolb, M., Paegelow, M., et al., 2014. Inductive pattern-based land use/cover change models: a comparison of four software packages. Environ. Modell. Software 51, 94–111.

Memarian, H., Balasundram, S.K., Talib, J.B., et al., 2012. Validation of CA-Markov for simulation of land use and cover change in the Langat Basin, Malaysia. J. Geogr. Inform. Syst. 4, 542–554.

Mishra, V.N., Rai, P.K., 2016. A remote sensing aided multi-layer perceptron-Markov chain analysis for land use and land cover change prediction in Patna district (Bihar), India. Arab. J. Geosci. 9 (4), 1–18.

Moulds, S., Buytaert, W., Mijic, A., 2015. An open and extensible framework for spatially explicit land use change modelling: the lulcc R package. Geosci. Model Develop. 3215–3229.

National Research Council, 2014. Advancing Land Change Modeling: Opportunities and Research Requirements. The National Academies Press, Washington, DC.

Paegelow, M., Camacho Olmedo, M.T., Mas, J., et al., 2013. Land change modelling: moving beyond projections. Int. J. Geogr. Inform. Sci. 27, 1691–1695.

Paegelow, M., Camacho Olmedo, M.T., Mas, J., et al., 2014. Benchmarking of LUCC modelling tools by various validation techniques and error analysis. Cybergeo – Eur. J. Geogr Systèmes, Modélisation, Géostatistiques.

Parsa, V.A., Yavari, A., Nejadi, A., 2016. Spatio-temporal analysis of land use/land cover pattern changes in Arasbaran Biosphere Reserve: Iran. Model. Earth Syst. Environ. 2, 178.

Perica, S., Foufoula-Georgiou, E., 1996. Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. J. Geophys. Res. Atmos. 101, 26347–26361.

Pontius Jr., R.G., 2018. Criteria to confirm models that simulate deforestation and carbon disturbance. Land 7 (3), 1–14.

Pontius Jr, R.G., Boersma, W., Castella, J., et al., 2008. Comparing the input, output, and validation maps for several models of land change. Ann. Reg. Sci. 42, 11–37.

Pontius Jr., R.G., Castella, J., De Nijs, T., et al., 2018. Lessons and challenges in land change modeling derived from synthesis of cross-case comparisons. In: Behnisch, M., Meine, G. (Eds.), Trends in Spatial Analysis and Modelling, Geotechnologies and the Environment, 19 edn. Springer International Publishing, Cham, Germany, pp. 143–164.

Pontius Jr., R.G., Gao, Y., Giner, N.M., et al., 2013. Design and interpretation of intensity analysis illustrated by land change in Central Kalimantan, Indonesia. Land 2, 351–369.

Pontius, R.G., Malanson, J., 2005. Comparison of the structure and accuracy of two land change models. Int. J. Geogr. Inform. Sci. 19 (2), 243–265.

Pontius Jr., R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int. J. Remote Sens. 32, 4407–4429.

Pontius Jr., R.G., Peethambaram, S., Castella, J., 2011. Comparison of three maps at multiple resolutions: a case study of land change simulation in Cho Don District, Vietnam. Ann. Assoc. Am. Geogr. 101, 45–62.

Quan, B., Ren, H., Pontius Jr., R.G., Liu, P., 2019. Quantifying spatiotemporal patterns concerning land change in Changsha, China. Landscape Ecol. Eng. 14 (2), 257–267. https://doi.org/10.1007/s11355-018-0349-y.

Raphael John, L., Hambati, H., Ato Armah, F., 2014. An Intensity Analysis of land-use and land-cover change in Karatu District, Tanzania: community perceptions and coping strategies. Afr. Geogr. Rev. 33, 150–173.

R Core Team, 2017. R: A Language and Environment for Statistical Computing.

Rocha, W.P., Barraza, G.C., Castro, G.G.H., et al., 2017. Spatial-temporal analysis of territorial transformations in the state of Sinaloa Mexico using geographic information systems. Agric. Sci. 8, 171–182.

Sang, L., Zhang, C., Yang, J., et al., 2011. Simulation of land use spatial pattern of towns and villages based on CA–Markov model. Math. Comput. Modell. 54, 938–943.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. John Wiley and Sons Ltd.

Shafizadeh-Moghadam, H., Asghari, A., Tayyebi, A., et al., 2017. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. Comput. Environ. Urban Syst. 64, 297–308.

Silva, E.A., Clarke, K.C., 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. Comput. Environ. Urban Syst. 26, 525–552.

Singh, S.K., Srivastava, P.K., Szabó, S., et al., 2015. Predicting spatial and decadal LULC changes through cellular automata markov chain models using earth observation datasets and geo-information. Environ. Process. 2, 61–78.

Sipper, M., 1997. Evolving uniform and non-uniform cellular automata networks. In: Stauffer, D. (Ed.), Annual Reviews of Computational Physics, 5th edn. World Scientific, Singapore, pp. 243–285.

Subedi, P., Subedi, K., Thapa, B., 2013. Application of a hybrid cellular automaton – Markov (CA-Markov) model in land-use change prediction: a case study of Saddle Creek Drainage Basin, Florida. Appl. Ecol. Environ. Sci. 1, 126–132.

Szabó, S., Novák, T., Elek, Z., 2012. Distance models in ecological network management: a case study of patch connectivity in a grassland network. J. Nature Conserv. 20, 293–300.

Túri, Z., 2010. Studying landscape pattern in Great Hungarian Plain model areas. In: Rahmonov, O. (Ed.), Anthropogenic Aspects of Landscape Transformations. University of Silesia, Sosnowiec-Będzin, pp. 109–115.

Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., Mastura, S.S.A., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. Environ. Manage. 30 (3), 391–405.

Veldkamp, A., Verburg, P.H., 2004. Modelling land use change and environmental impact. J. Environ. Manage. 72, 1–3.

Yang, Y., Liu, Y., Xu, D., et al., 2017. Use of intensity analysis to measure land use changes from 1932 to 2005 in Zhenlai County, Northeast China. Chin. Geogr. Sci. 27 (3), 441–455.

Yang, X., Zheng, X., Chen, R., 2014. A land use change model: integrating landscape pattern indexes and Markov-CA. Ecol. Model. 283, 1–7.

Zhou, P., Huang, J., Pontius Jr., R.G., et al., 2014. Land classification and change intensity analysis in a coastal watershed of Southeast China. Sensors 14 (7), 11640–11658.

Ziółkowska, E., Ostapowicz, K., Radeloff, V.C., et al., 2014. Effects of different matrix representations and connectivity measures on habitat network assessments. Landscape Ecol. 29, 1551–1570.